AD-A099 395 MASSACHUSETTS INST OF TECH LEXINGTON LINCOLN LAB F/6 17/2
FRAME-FILL TECHNIQUES FOR REDUCING VOCODER DATA RATES.(U)
FEB 81 P E BLANKENSHIP, M L MALPASS F19628-80-C-0002
UNCLASSIFIED TR-556 ESD-TR-81-39 NL

1 OF 1
AD A
099395

END
DATE
FILMED
6-81
DTIC

LEVEL II

12

AD A099395

Technical Report 556

P.E. Blankenship
M.L. Malpass

DTIC
SELECTE
MAY 27 1981
S C

Frame-Fill Techniques for Reducing Vocoder Data Rates

26 February 1981

Prepared for the Department of the Air Force
under Electronic Systems Division Contract F19628-80-C-0002 by

Lincoln Laboratory
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Lexington, Massachusetts

Approved for public release; distribution unlimited.

The work reported in this document was performed at Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology, with the support of the Department of the Air Force under Contract F19628-80-C-0002.

This report may be reproduced to satisfy needs of U.S. Government agencies.

The views and conclusions contained in this document are those of the contractor and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the United States Government.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

Raymond L. Loiselle

Raymond L. Loiselle, Lt.Col., USAF
Chief, ESD Lincoln Laboratory Project Office

Non-Lincoln Recipients

PLEASE DO NOT RETURN

Permission is given to destroy this document when it is no longer needed.

12

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

LINCOLN LABORATORY

FRAME-FILL TECHNIQUES FOR REDUCING VOCODER DATA RATES

P.E. BLANKENSHIP
M.L. MALPASS
Group 24

DTIC
ELECTE
MAY 27 1981
D
C

TECHNICAL REPORT 556

26 FEBRUARY 1981

Approved for public release; distribution unlimited.

LEXINGTON MASSACHUSETTS

ABSTRACT

Two experimental vocoder systems are described which exploit the frame-fill concept described by McLarnon to achieve data rates in the range of 800 to 1200 bps. One is based on a well-known 2400 bps channel vocoder design, the second is based on a form of the Lincoln 2400 bps linear predictive coder (LPC-10) algorithm. Both systems were found to perform well at the 1200 bps rate representing a 2:1 savings in transmission bandwidth at very little additional algorithm complexity. At 800 bps both systems were judged usable but not wholly satisfactory. Performance of the channel vocoder was considered marginally better than LPC at 800 bps.

Accession For
NTIS GRA&I
DTIC TAB
Unannounced
Justification
By
Distribution/
Availability Codes
Avail and/or
Dist Special
A

CONTENTS

I. INTRODUCTION

Planners of advanced future communications systems continue to challenge voice digitization system designers with demands for further reductions in required transmission bandwidth. Given the practical balance which must be struck between algorithm performance and implementational complexity, 2400 bits/sec represents a reasonable lower limit on the capability of speech bandwidth compression technology today. Even at this rate, performance is still a major issue within the military operational community where it is frequently necessary to accommodate acoustically noisy and heavily jammed environments.

Several approaches have been or are presently being pursued in an effort to develop voice digitization algorithms supporting data rates well below 2400 bps. Based on formant tracking, vector quantization, pattern matching, or diphone analysis concepts, these systems are still highly embryonic, experimental, and in some cases quite computationally complex.

This report describes a near-term, low-complexity, low-risk approach to achieving data rates in the 800 to 1200 bps range based on standard 2400 bps analysis-synthesis system technology. Using conventional filter bank-based or linear prediction (LPC) type systems as fundamental backbones, it is shown that surprising performance is possible in this bit rate range with essentially negligible increases in implementational complexity.

II. THE FRAME-FILL CONCEPT

The frame-fill concept described by McLarnon [1] represents a conceptually simple and straightforward approach to reducing the transmission bandwidth requirement of a frame-oriented digital voice system. The basic idea is to transmit from analyzer to synthesizer only every Mth data frame, thereby achieving an approximate M:1 reduction in rate. The savings is only approximate in that some control information must also be supplied to the receiver instructing the synthesizer how to reconstruct (or "fill in") missing information given some pre-agreed fixed set of options.

This idea is of great interest when applied to 2400 bps vocoder systems which represent a practical lower bound on the data rate capability of today's voice digitization algorithm technology. By choosing M = 2, a 1200 bps data rate would be achieved. If M = 3, an 800 bps system would appear feasible. When the mechanics of human speech production and perception are taken into consideration, M = 2 is the most reasonable choice if starting with a 2400 bps system. This stems from the fact that narrowband vocoders typically operate at fundamental frame production rates of about 50 Hz. This is close to a practical minimum if essential phonemic transitions in the speech are to be reasonably preserved. If more than 50% of the vocoder analysis frames are omitted, it is effectively impossible to avoid unacceptable losses in this vital transitional information which bears so significantly on intelligibility.

This report will focus its attention on nominal 2400 bps systems with

M = 2; i.e., every other frame produced by the vocoder analyzer will be omitted from the transmitted data stream. Given this constraint, the following general set of rules will be applied in determining the best "fill in" option at the vocoder synthesizer:

(1) Compare the frame of data to be omitted with the frames immediately preceding and succeeding in the temporal sense.

(2) In accordance with some "reasonable" distance metric, decide which neighbor matches the frame to be omitted most closely.

(3) Also consider as a match candidate some weighted combination of the information contained in the two neighboring frames.

(4) Select the option (3 choices) representing the best match and append its I.D. code (2 extra bits) to the frame which is to be transmitted.

In practice these rules apply most naturally to the vocal tract parametric information. In a channel vocoder this corresponds to the spectral samples, in an LPC vocoder the K-parameter set. Also, the 3rd fill-in option is usually constrained to be a simple average of the neighboring frame data. The excitation information, however, is handled separately.

Both systems discussed here treat the pitch and voicing parameters in the same, empirically determined way. It is known that the voicing information plays a critical role in determining the quality and intelligibility of the synthetic speech. Since it comprises only a single bit per frame, the penalty for transmitting <u>every</u> frame time is minimal.

Given that all the voicing information is available at the receiver, intelligent decisions can be made on reconstructing the omitted pitch parameter (if needed) from information in the neighboring frames. Of course, if the omitted frame happens to be unvoiced, then a pitch parameter may not be needed.

The pitch/voicing reconstitution strategy is summarized in Table I where the excitation parameter is developed based on the 8 possible combinations of voicing bits that might be encountered. Frame "N" is to be reconstituted, and only its voicing bit is available to the receiver. Implicit in this strategy is some editing of the voicing decision itself aimed at rejecting improbable combinations. The pitch fill-in approach suggested appears from our experience to offer the most favorable perceptual impact.

Given this common approach to dealing with the excitation information, the next two sections will focus on methods for reconstruction of the vocal tract data unique to channel and LPC types of 2400 bps backbone vocoder.

TABLE I

PITCH/VOICING RECONSTITUTION

FRAME			EXCITATION
N - 1	N	N + 1	
UV	UV	UV	UV
UV	UV	V	UV
UV	V	UV	UV
UV	V	V	P_{N+1}
V	UV	UV	UV
V	UV	V	$1/2[P_{N+1} + P_{N-1}]$
V	V	UV	P_{N-1}
V	V	V	$1/2[P_{N+1} + P_{N-1}]$

V = VOICED

UV = UNVOICED

P = PITCH PARAMETER

III. FILTER BANK-BASED SYSTEM

The channel vocoder system used as the basis for this series of experiments was modelled after the UK JSRU vocoder [2] sometimes referred to as the "Belgard" algorithm. This vocoder transmits 19 spectral samples per 20 msec spanning the frequency range of 200 to 4000 Hz and features relatively simple analyzer and synthesizer bandpass filter designs.

In this framework the vocal tract parametric information assumes the form of 19 logarithmically uncoded spectral samples. McLarnon suggested that the 3rd fill-in choice be constructed by averaging the log spectral data on a channel-by-channel basis. He further suggested a very simple distance metric of the form

$$\varepsilon_1 = \sum_{k=1}^{19} |\tilde{S}_c(k) - \tilde{S}_r(k)| \tag{1}$$

where $\tilde{S}_c(k)$ and $\tilde{S}_r(k)$ are the kth log spectral samples of the candidate and reference frames respectively. In other words, this is simply the sum over all the vocoder channels of magnitudes of the differences in the spectral samples. This simple metric offers the obvious advantage of being very easily computed.

In the course of experimentation at least three other metrics of similar complexity were tried as summarized below.

$$\varepsilon_2 = \sum_{k=1}^{19} |\tilde{S}_c(k) - \tilde{S}_r(k)| \cdot \tilde{S}_r(k) \tag{2}$$

$$\varepsilon_3 = \underset{k}{\mathrm{Max}} \left| \tilde{S}_c(k) - \tilde{S}_r(k) \right| \tag{3}$$

$$\varepsilon_4 = \sum_{k=1}^{18} \left| \{\tilde{S}_c(k+1) - \tilde{S}_c(k)\} - \{\tilde{S}_r(k+1) - \tilde{S}_r(k)\} \right| \tag{4}$$

These were developed in the hope of finding one offering a more favorable overall perceptual impact. Metric (2) represents a weighted version of (1) where a given spectral difference influences the overall metric in a manner consistent with the spectral amplitude at that point in frequency. Metric (3) seeks to minimize the maximum single point error over the 3 fill-in alternatives. Metric (4), suggested by Klatt [3], is designed to emphasize differences in spectral slope.

Experience indicated little performance difference among (1), (2), and (3) with a slight preference, based on informal listening, emerging for (2). Metric (4), however, was found to be decidedly inferior. A possible reason is failure to take into account total spectral energy where this information is implicit to some extent in each of the other metrics. It is therefore possible that two spectra with wildly differing energy content could be equated because of similarities in formant structure. Metric (4) by itself was therefore not considered satisfactory for present purposes.

Both 1200 and 800 bps versions of the channel vocoder system were developed as summarized in Table II. The 1200 bps system is based on the 2400 bps system as shown. Here 48 bits are transmitted per 20 msec and 2 are normally uncommitted. A 2-bit-per-channel DPCM type of coding

TABLE II

CODING CONVENTIONS FOR CHANNEL VOCODER-BASED SYSTEM

SYSTEM (BPS)	VOICING	PITCH	REF. CHANNEL	2-BIT DPCM	1-BIT DPCM	CTRL	TOTAL BITS	FRAME RATE (HZ)
2400	1	6	3	18	0	0	48*	50
1200	2	6	3	17	1	2	48	25
800(V)	2	6	3	1	17	4	33	24.24
800(UV)	2	0	3	6	12	4	33	24.24

*2 BITS UNUSED

scheme is used to represent the channel weights. The lowest frequency channel (240 Hz) is used as a starting point and is coded in a 3-bit log PCM format. The 1200 bps variant transmits 48 bits each 40 msec. The two formerly uncommitted bits are used to convey the spectral fill-in option dictated by the analyzer. The highest frequency channel is represented as a 1-bit DPCM datum thereby freeing a bit to represent the voicing state of the omitted frame. The transmitted frame, therefore, contains 6 pitch bits (log coded), 2 voicing bits, 2 fill-in control bits, and 38 spectrum bits.

The 800 bps version is also a 2:1 reduction system and therefore starts with a 1600 bps backbone. If all of the spectral information is to be included, the minimum number of spectral bits possible (relying on 1-bit DPCM coding) is 3 + 18 = 21. Given the remaining information that must be included (6 + 2 + 2 = 10), the net minimum for a fill-in system would be 31 bits/frame. At 20 msec/frame, this results in a net data rate figure of 1550 bps.

The 1-bit DPCM coding allows step sizes normally fixed at $\pm$ 6 dB, a reasonable compromise in most cases. However, the 1-bit DPCM coding does a significantly poorer job of representing the spectrum than 2-bit DPCM. An effort was made to improve the quality of the 1600 bps spectral representation by permitting some flexibility in the <u>choice</u> of step size for the 1-bit DPCM coding. A system was developed which permitted step size choices of 3, 4.5, 6, and 7.5 dB. The step size choice was determined by comparing the coded spectrum with the uncoded reference for each of

the 4 options. A distance metric similar to the one used for frame fill-in purposes was used as the determinant. Once the best fit was determined, then the normal 2:1 frame rate reduction process was put into effect. Again metric (2) was found to be the most useful. Given four choices of step size, 2 more control bits were appended to the transmit frame for a total of 33. To reduce the data rate to exactly 1600 bps, the frame period was lengthened slightly to 20.625 msec.

In considering ways to further improve quality, it was realized that the 6 pitch bits were being effectively wasted during unvoiced frames. It was decided to give these bits over to the spectral representation during unvoiced frames resulting in a combined 1-bit/2-bit DPCM spectral coding where channels 2 through 7 received the 2-bit accuracy. Although in benign environments spectral detail is not as critical in consonantal sounds as it is during vowels, there is some evidence that it is of importance in noisy backgrounds. The resulting 800 bps system is summarized at the bottom of Table II.

The 1200 bps system was found to perform quite well based on informal listening tests using high quality (acoustically quiet background, dynamic microphone) input material. In many instances the 1200 bps output could barely be distinguished from the 2400 bps parent. The 800 bps system was observed to do surprisingly well for its rate and complexity but was generally judged markedly inferior to the 1200 bps version. Under relatively benign conditions it is probably usable, especially in the hands of properly trained personnel. Under the degraded conditions typical of many military environments it would probably not be acceptable.

IV. LINEAR PREDICTION-BASED SYSTEM

Low rate systems based on Linear Predictive Coder (LPC) backbones are of particular interest within the DoD community since this class of algorithm has been selected as the standard for 2400 bps applications, and all interoperability criteria are based on a particular LPC formulation. The experimental system described here is based on the Lincoln 10th order autocorrelation LPC [4] modified to be consistent with the essential elements of the DoD interoperability specification [5,6]. Modifications include 22.5 msec non-overlapped framing, digital audio conditioning, a 4th order LPC fit during unvoiced frames, and implementations of NSA-specified coding tables for pitch/voicing, energy, and K-parameters.

Although the philosophical approach taken is identical to that employed in the channel vocoder, the parametric information developed in an LPC system is unique and must be given special consideration. For example, the vocal tract is represented by a 10th order, all-pole digital filter. The filter is characterized for transmission purposes in terms of 10 K-parameters which can be applied directly at the synthesizer if the synthesis filter is implemented in a lattice ("acoustic tube") form. Alternatively, the filter could be implemented in a direct form requiring the so-called direct form ("a") parameters.

There are many possible distance metrics that have been discussed in the literature which could be developed from these or other parametric representations. In a study by Barnwell [7] metrics of the form shown below were evaluated in terms of perceptual impact on human subjects.

$$\varepsilon = \sum_{j=1}^{P} |Q_c(j) - Q_r(j)|^N \qquad (5)$$

Possible choices of parameter sets included a-parameters, K-parameters, area ratios, and log area ratios. It was found that the choice of K-parameters with $N \sim 1$ gave good correlation with perceptual feedback. That is, the parameter set that minimizes ε when based on the sum-of-magnitudes differences in K-parameter space also <u>sounds</u> the closest to the reference set. This fortunate result leads to a metric for LPC systems which has identical form to that of the channel vocoder:

$$\varepsilon_{LPC} = \sum_{j=1}^{10} |K_c(j) - K_r(j)| \qquad (6)$$

In the spirit of the channel vocoder, the fill-in options permitted were: fill forward, fill backward, or fill with an average of the K-parameter sets.

However, the LPC formulation includes a separate parameter which takes into account total spectral energy content. In the experimental system the energy parameter was treated separately and was reconstructed independently of the spectral data as shown below:

$$\varepsilon_{RMS} = |\log E_c - \log E_r| \qquad (7)$$

where the fill-in options were constrained to be fill forward, fill backward, or fill with the average of the log energies.

All of these metrics feature the desirable property of extreme computational simplicity, and there is no need to develop intermediate functional representations (e.g., Itakura-Saito metric). Also implicit is the requirement for 2 sets of fill-in control bits since energy and spectrum are treated separately.

Both 1200 and 800 bps systems were developed and evaluated as summarized in Table III. Notice that pitch and voicing in the 1200 bps version are handled exactly as they were in the channel vocoder. Four fill-in control bits are supplied to decouple energy from spectrum. The parameter coding strategy shown in the table represents a necessary departure from the DoD standard [5] due to the extra 5 bits required for fill-in and voicing control. K3 has been reduced from 5 to 4 bits, K6 from 4 to 3, K7 from 4 to 2, and K8 from 3 to 2. These choices evolved empirically through informal listening tests conducted with trained personnel.

The 800 bps system was much more difficult to implement given the lack of a truly efficient parameter coding scheme akin to the 1-bit DPCM technique applied in the channel vocoder. To create the necessary 1600 bps backbone, 18 out of 54 bits per frame must be discarded. To accomplish this K5 through K9 were reduced to 1 bit each, energy and KØ through K2 were reduced to 4 bits each, and K3 and K4 were dropped to 3 and 2 bits respectively. This was judged to be the maximum that could be stripped from the spectral representation if 10 K-parameters were to be retained. To gain the remaining number of bits required, the explicit sync bit was

TABLE III

CODING CONVENTIONS FOR LPC-BASED LOW RATE SYSTEMS

	1200 BPS	800 BPS	
SYNC	1	0	
V/UV THIS FRAME	1	0	
V/UV NEXT FRAME	1	1	
STRATEGY BITS	4	3	
PITCH POINTER	6	6	
ENERGY	5	4	
K0	5	4	
K1	5	4	
K2	5	4	
K3	4	3	
K4	4	2	OMITTED IN UNVOICED FRAMES
K5	4	1	
K6	3	1	
K7	2	1	
K8	2	1	
K9	2	1	

dropped, the present-frame voicing bit was packed into the 6-bit pitch word (by sacrificing one pitch code), and the number of control bits was reduced from 4 to 3. The reduction in control bits was accomplished by noting that 4 bits were being used to represent 3 x 3 = 9 combinations of spectral and energy fill-in choice. Statistics were gathered over a large body of speech data indicating that one of the 9 combinations occurred rather infrequently. This case, where encountered, was automatically mapped into one of the remaining 8 choices. With only 8 permitted combinations, 3 control bits suffice.

As in the case of the channel vocoder, informal listening tests indicated that the 1200 bps LPC performed nearly as well the 2400 bps version. However, the 800 bps version was judged to be inferior to its channel vocoder counterpart. This is probably due to the fact that a crude-but-efficient spectral coding scheme analogous to 1-bit DPCM is not presently known for LPC parameters.

V. FORMAL PERFORMANCE EVALUATIONS

Both the filter bank and LPC-based systems were subjected to formal evaluation through the diagnostic rhyme test (DRT).* Each system was tested at rates of 2400, 1200, and 800 bps. The source material was comprised of 3 male talkers in an acoustically quiet background using a high-quality dynamic microphone. The results are summarized in Table IV.

Based on these results alone, it appears as though both 1200 and 800 bps data rates are usable in relatively benign environments although some informal in-house communicability tests tended to refute this conclusion at 800 bps. Note that the LPC-based systems scored slightly below the channel vocoder at 1200 bps and 800 bps which is in agreement with impressions gained through informal listening. As stated previously, this is probably due to the lack of a truly efficient parameter coding scheme for LPC equivalent to the DPCM spectral coding approach used in the channel vocoder.

*DRT scoring services provided by RADC/EEV speech laboratory.

TABLE IV

SUMMARY OF 3-SPEAKER DRT RESULTS

RATE (BPS)	CHANNEL VOCODER	LPC VOCODER
2400	89.6	91.6
1200	87.6	85.1
800	84.0	82.0

VI. RATE COMPATIBILITY AND EMBEDDED CODING

The reduced rate systems described evolve as simple extensions of their respective 2400 bps parents and are parameter-compatible with them. Given this close relationship, it is reasonable to consider whether or not the frame-fill approach is consistent with so-called embedded coding concepts and whether high and low rate versions of the same system can be made to intercommunicate in some reasonable (though probably suboptimum) way. Given the freedom to design properly transmission protocols and formats a priori, the answer would appear to be positive as discussed below.

The concept of embedded coding makes the most sense in the context of an intelligent connectivity medium such as a packetized digital multifunction communications network [8]. The ARPANET and JTIDS are examples of this type of advanced and sophisticated system. The basic idea is to produce a high rate data stream at the transmission source in which is embedded one or several lower data rate streams. The network, having knowledge of which data is essential and which is expendable, can delete the less critical information according to some pre-agreed set of priorities if circumstances warrant. It might do this in response to fluctuations in available channel capacity caused by heavy traffic or severe jamming. In the case of a nominal 2400 bps voice digitizer, the network could cut the data rate effectively in half by invoking the frame-fill mode if a discretionary mechanism were available for it to do so.

The channel vocoder will be considered an example of how this might be accomplished. The transmitter could be modified to perform the frame-fill control computations on successive pairs of frames at <u>all</u> times and form a pair of packets as shown in Fig. 1. Packet P_1 is the high priority member of the pair and contains the essential information. It is comprised of the usual pitch (P_n) voicing (V_n), and spectral data (S_n). However, it is augmented to include the voicing bit from the contiguous frame (V_{n+1}) and a 2-bit control field (C_n) indicating how best to reconstruct the pitch and spectral data for the contiguous frame (cf. Section III). The actual pitch and spectral information for that frame along with miscellaneous unused bits are contained in the lower priority packet, PØ. The two packets together account for 96 bits which are transmitted in a 40 msec epoch. Thus the transmitter produces a constant 2400 bps data stream from which appropriately designated information can be stripped at will reducing the net data rate to 1200 bps (for this particular format convention).

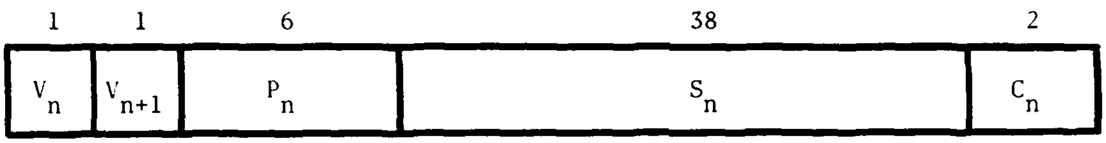
It remains for the network to notify the receiver when the data rate has been so modified, which is an easy task for a medium of this presumed type. The receiver will then invoke its frame-fill logic to operate on the data actually received in accordance with the techniques described in Section III. If the data rate has not been modified, the receiver will ignore the frame-fill control data and operate normally at the 2400 bps rate.

It is also interesting to consider the possibility of rate compatibility

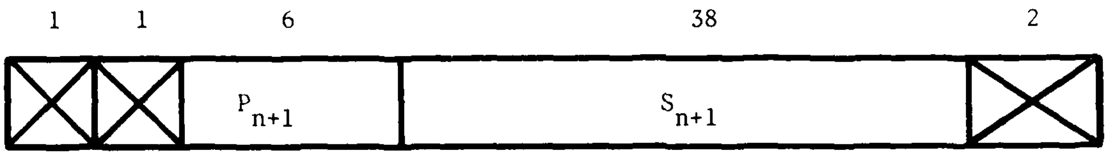
between 1200 and 2400 bps variants of the same generic vocoder type given unsophisticated connectivity. It would be desirable, for example, for a 1200 bps source to be able to communicate with both 1200 and 2400 bps receivers without any specific knowledge of which type might actually be at the other end of the link. Conversely, it might be useful to have a 1200 bps receiver which can absorb data from both 1200 and 2400 bps sources without any special control interactions.

Considering again the channel vocoder, assume that the transmitter whether in a 1200 or 2400 bps mode produces frames of data formatted as shown in Fig. 1(a). If operating at 2400 bps, one frame will be transmitted each 20 msec; at 1200 bps the transmission epoch will be 40 msec. Also assume that in the 2400 bps mode the bit stream is arranged to interlace bits on a frame pair basis as indicated in Fig. 2. This arrangement insures that each bit of frame n is immediately succeeded in the transmission stream by the corresponding bit of the contiguous frame (n + 1). Implicit in this strategy is that each voicing bit is transmitted twice and that synchronization is based on a frame pair (96 bits). In the 1200 bps mode, there is no interlacing.

It is necessary now to consider the four possible situations that could occur: 2400→2400, 2400→1200, 1200→2400, 1200→1200. If the transmitter and receiver rates are matched, there is no problem. If a 2400 bps source is being received by a 1200 bps sink, only every other bit will be received. The interleaved format guarantees that every other frame will be absorbed in its entirety. Since frame-fill data is present



(a). Priority P1 packet (48 bits).



(b). Priority PØ packet (48 bits).

Fig. 1(a-b). Prioritized packets for embedding a channel vocoder.

$$V_n V_{n+1} V_{n+1} V_{n+2} P5_n P5_{n+1} \cdots P0_n P0_{n+1} S37_n S37_{n+1} \cdots\cdots C0_n C0_{n+1}$$

Fig. 2. Bit interleaving within a frame pair.

in <u>every</u> frame, no special synchronization strategy is necessary and the 1200 bps synthesizer can function in its usual way.

If a 1200 bps source is being received by a 2400 bps sink, it will clock in each received bit twice. Since it assumes an interlaced format, it will de-multiplex two identical frames and synthesize accordingly. Speech quality will be considerably poorer than it would be with the frame-fill mechanism operative, but the link will probably be usable. On the other hand, the receiver could be equipped with frame-fill logic and some means could be provided for it to determine trivially that it's connected to a 1200 bps stream. This might be accomplished by noting that successive bits are always pair-wise identical, or a special bit in the transmission format could be provided as a rate ID.

The principles described above could be applied equally well to an LPC-10 type of vocoder. However, the transmission format presently specified in the DoD narrowband interoperability standard is not appropriate for supporting this kind of flexibility although there are no conceptual problems with the vocoder algorithm per se. A modified format would be necessary if these compatibility features were deemed essential.

VII. SUMMARY AND CONCLUSIONS

Methods have been described based on the principle of frame fill-in for developing reduced rate transmission systems from standard 2400 bps backbones. It was shown that both channel vocoder and LPC types of vocoders could be adapted with virtually no increase in computational complexity to operate at 1200 or 800 bps. The compatibility of this approach with embedded coding concepts was discussed.

It was found through informal and formal evaluation methods that both channel vocoder and LPC-based systems perform quite well at 1200 bps and would probably be usable in most environments where the 2400 bps parent could be successfully operated. At 800 bps both systems were considered marginal and usable only in limited circumstances. However, the channel vocoder was seen to perform incrementally better, probably due to the uniquely efficient parameter coding scheme employed which tends to be less sensitive to quantization inaccuracies.

REFERENCES

1. E. McLarnon, "A Method for Reducing the Frame Rate of a Channel Vocoder by Using Frame Interpolation," ICASSP '78, Washington, D.C., pp. 458-461 (April 1978).

2. J. Holmes, "The JSRU Channel Vocoder," IEE Proc. on Communications, Radar, and Signal Processing, Vol. 127, Part F, No. 1, pp. 53-60 (February 1980).

3. D. Klatt, private communication.

4. P. Blankenship et al., "The Lincoln Digital Voice Terminal System," Technical Note 1975-53, Lincoln Laboratory, M.I.T. (25 August 1975), DDC AD-A017569/5.

5. T. Tremain et al., "Implementation of Two Real Time Narrowband Speech Processing Algorithms," IEEE Elect. and Aerospace Systems Convention Record, pp. 698-708 (September 1978).

6. M. Malpass, private communication.

7. T. P. Barnwell and W. D. Voiers, "An Analysis of Objective Measures for User Acceptability of Voice Communications Systems," Final Report, DCA100-78-C-0003 (September 1979).

8. T. Bially, B. Gold, and S. Seneff, "A Technique for Adaptive Voice Flow Control in Integrated Packet Networks," IEEE Trans. Commun., COM-28, 325 (1980).

VII. SUMMARY AND CONCLUSIONS

Methods have been described based on the principle of frame fill-in for developing reduced rate transmission systems from standard 2400 bps backbones. It was shown that both channel vocoder and LPC types of vocoders could be adapted with virtually no increase in computational complexity to operate at 1200 or 800 bps. The compatibility of this approach with embedded coding concepts was discussed.

It was found through informal and formal evaluation methods that both channel vocoder and LPC-based systems perform quite well at 1200 bps and would probably be usable in most environments where the 2400 bps parent could be successfully operated. At 800 bps both systems were considered marginal and usable only in limited circumstances. However, the channel vocoder was seen to perform incrementally better, probably due to the uniquely efficient parameter coding scheme employed which tends to be less sensitive to quantization inaccuracies.

(14) TR-556

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(11) **REPORT DOCUMENTATION PAGE**

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER (18) ESD-TR-81-39	2. GOVT ACCESSION NO. AD-A099395	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Frame-Fill Techniques for Reducing Vocoder Data Rates ,		5. TYPE OF REPORT & PERIOD COVERED (9) Technical Rep[redacted]rt,
		6. PERFORMING ORG. REPORT NUMBER Technical Report 556
7. AUTHOR(s) Peter E. Blankenship [redacted] Marilyn L. Malpass		8. CONTRACT OR GRANT NUMBER(s) (15) F19628-80-C-0002
9. PERFORMING ORGANIZATION NAME AND ADDRESS Lincoln Laboratory, M.I.T. P.O. Box 73 Lexington, MA 02173		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Program Element No. 33401F Project No. 7820
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Systems Command, USAF Andrews AFB Washington, DC 20331		(11) 12. REPORT DATE 26 Febr[redacted]81
		13. NUMBER OF PAGES 32
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Electronic Systems Division Hanscom AFB Bedford, MA 01731 (12)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

None

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

speech compression	embedded coding
channel vocoder	frame fill
linear predictive coder	speech bandwidth reduction
packet speech	low rate vocoder

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Two experimental vocoder systems are described which exploit the frame-fill concept described by McLarnon to achieve data rates in the range of 800 to 1200 bps. One is based on a well-known 2400 bps channel vocoder design, the second is based on a form of the Lincoln 2400 bps linear predictive coder (LPC-10) algorithm. Both systems were found to perform well at the 1200 bps rate representing a 2:1 savings in transmission bandwidth at very little additional algorithm complexity. At 800 bps both systems were judged usable but not wholly satisfactory. Performance of the channel vocoder was considered marginally better than LPC at 800 bps.

DD FORM 1473 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

207650